

Defining and Evaluating Context for Wearable Computers

Bristow, Huw; Baber, Christopher; Cross, James; Knight, James; Woolley, Sandra

DOI:

[10.1016/j.ijhcs.2003.11.009](https://doi.org/10.1016/j.ijhcs.2003.11.009)

Document Version

Early version, also known as pre-print

Citation for published version (Harvard):

Bristow, H, Baber, C, Cross, J, Knight, J & Woolley, S 2004, 'Defining and Evaluating Context for Wearable Computers', *International Journal of Human-Computer Studies*, vol. 60, pp. 798-819.
<https://doi.org/10.1016/j.ijhcs.2003.11.009>

[Link to publication on Research at Birmingham portal](#)

General rights

Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

- Users may freely distribute the URL that is used to identify this publication.
- Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
- User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
- Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

Take down policy

While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact UBIRA@lists.bham.ac.uk providing details and we will remove access to the work immediately and investigate.

Defining and evaluating context for wearable computing

Huw W. Bristow, Chris Baber*, James Cross, James F. Knight,
Sandra I. Woolley

*Kodak/Royal Academy Educational Technicon, School of Electronic, Electrical and Computer Engineering,
The University of Birmingham, Edgbaston, Birmingham B15 2TT, UK*

Received 17 November 2003; accepted 21 November 2003

Abstract

Defining ‘context’ has proved to be a non-trivial problem for research in context-awareness. In this paper we address two questions: what features of activity are required to define context? and does the use of context-awareness measurably improve user performance? The first question was addressed by a study of everyday activities, using a Photo Diary method to arrive at a set of Context Identifiers. We feel that it is important to discover what features of activity are needed in order to describe context. Two user trials were carried out to address the second question. We conclude that the use of context improves user task proficiency.

© 2004 Elsevier Ltd. All rights reserved.

1. Introduction

Recent developments in wearable and mobile computing are based on the assumption that ‘context’ can be related to concepts such as the user’s location and activity. Ostensibly this could allow associative structures (or ‘models’) that link features of context to items of information; as the features of context change, so too will the relevance of information. Thus, a significant aspect of wearable and mobile computing is the ability of the technology to respond to changes in ‘context’. Unfortunately, there remains a distinct lack of agreement as to what constitutes ‘context’. Abowd and Mynatt (2002) suggest that one can think of context in terms of: who is using the system; what the system is being used for; where the system is

*Corresponding author. Tel.: +44-121-414-3965; fax: +44-121-414-4291.

E-mail addresses: huw@huwbristow.co.uk (H.W. Bristow), c.baber@bham.ac.uk (C. Baber).

being used; when the system is being used and why the system is being used. This provides an initial avenue into the problem of defining context. We suggested a classification scheme (Baber et al., 1999), as illustrated in Table 1, which pairs a Reference Marker, i.e. the element that is being defined, with a simple demarcation of time. In this manner context can be defined in terms of a combination of Reference Markers that have different information relating to whether the information is stored, whether it is predicted, or whether it is being captured at the moment.

Table 1 draws upon the example of a visit to an art gallery. The visitor carries a mobile telephone and receives ‘text alerts’ (perhaps from the gallery itself to describe events that are taking place, say a lecture in one of the galleries). The visitor’s primary goal is to look at the paintings and to learn something about them. The ‘current time’ column indicates how this is achieved. However, the system is assumed to also make reference to data from ‘past’ events, settings, activities, etc. and draw upon these data to construct the information provided to the visitor, and to also make predictions of possible actions. A question of prime importance, therefore, is what data ought to be used to construct the ‘context’ for such a system?

An early example of a context-aware application, Forget-Me-Not (Lamming and Flynn, 1994), sought to offer a ‘memory prosthesis’ by collecting information about specific events, such as where or when a meeting occurred or who was present at the meeting, in order to provide a reminder to the user in support of subsequent recall of the event. The Remembrance Agent (Rhodes, 1997) provided users with links to textual material, e.g. emails and reports, which was relevant to a specific event. More recent systems, such as Shopping Assistant (Asthana et al., 1994), ComMotion (Marmasse and Schmandt, 2000), and CyberMinder (Dey and Abowd, 2000) also provide the user with an application that can support specific activities, often supporting the recall of some information or providing a reminder to perform a task, based on the ‘context’ of the person. In other applications, context is used to select and present relevant information to the user. For example, in the ‘Touring Machine’ (Feiner et al., 1997), environmental reference markers using Global Positioning System (GPS), were used to link a person’s location with both pertinent information,

Table 1
Features of context

Reference markers	Past	Current time	Future
Event	Log of previous alerts	Incoming text alert	Schedule of alerts
Environment	Route taken to current location	Current location	Possible destinations from the current location
Task	Been Done list	Current activity	To Do list
Person	Acquired knowledge of art	Current interests; current physiological state	Possible things to discover
Object	Previously viewed paintings	Current painting	Possible similar paintings

e.g. a description of the building that they were near, and augmentation of their view of the world, i.e. through superimposing arrows on the head-up display to point to buildings.

Applications have seen context being defined by environmental characteristics, e.g. Schmidt et al. (1999) report a device (mounted on the wearer's tie) that can detect changes in ambient sound levels or wearer movement. Similarly there has been interest in the notion of sensors mounted in badges that could be used to detect a person's location, and can modify information relating to their environment accordingly (Want et al., 1992). Alternatively, context could be defined by changes in the physical and physiological state of the wearer (Farrington et al., 1999; Lind et al., 1997). Finally, context can be related to the task that the person is performing. Thus, wearable computers can assist the wearer in shopping activities (Asthana et al., 1994; Randell and Muller, 2000).

The definition of 'context' represents a challenge to researchers in this field (Baber, 2001). Sensibly, researchers have tended to focus on defining the application and seeking to measure or sense aspects of the context that are appropriate to the application. Thus, in their study of delivery van drivers, Ashbrook and Starner (2002) concentrated on GPSs and developed sophisticated means of collating these data into patterns of activity. On the one hand, the approach to context-awareness as the measurement of features is pragmatic, but on the other hand, it is not clear whether these features are the most appropriate; we are relying on the intuitions of the researchers to ensure that context is defined correctly. While such a 'craft-based' approach is proving fruitful in many applications, it tends to lead to researchers rediscovering (or at least redefining) the whole concept of context. This, in turn, is producing a degree of circularity into both the literature and the research community, i.e. 'context' is in danger in being defined simply in terms of whatever features a 'context-aware' device is able to measure. Furthermore, while it is taken on trust that context-awareness is a good thing, there has been surprisingly little by way of research into the potential benefits of context-awareness on user performance.

The first aim of this paper is to take an activity-centred approach to the definition of context. In doing so, we hope to explore the meaning of context for a host of everyday activities. From this we can begin to develop a comprehensive set of context identifiers. The second aim of the paper is to examine the potential benefits of context-aware devices, particularly in comparison with their counterparts that do not exhibit such awareness.

2. Study one

The goal of study one is to investigate the context of everyday activity. In order to achieve this goal, it was decided that participants would record everyday activity. In this study a variation of a diary study (Rieman, 1993) was employed. It was felt that asking participants to complete a written diary could prove difficult in that they might not be sure what to record (hence, leading to problems of 'limited recording' in

which participants recording only aspects of the activity that they find easy to put into words and producing only limited accounts) and that recording might be unduly intrusive. Furthermore, a diary study might also have led ‘recording bias’, i.e. to participants recording information that they felt was relevant to the study rather than examples of the mundane activities that we sought. Consequently, it was proposed that participants would record instances of their everyday activities using photographs. Eldridge et al. (1992) previously reported a study into the use of video-recording as a means of creating a record of everyday activity. Tolmie et al. (2002) and Brown et al. (2000) have reported the use photographic diary methods to study everyday activities. Brown et al. (2000) limited the activities studied to ‘information capture’ events in the work place and did not apply their work to the definition of context. Tolmie et al. (2002) looked at a set of ‘domestic routines’ (specifically interactions between mothers leaving a house to pick up their children from school) and discuss the context of these interactions in order to show how context plays an implicit role in understanding the interactions, although they do not apply their findings to the definition of context.

In this study, participants were asked to photograph everyday activities. However, in order to overcome potential recording bias, it was decided that taking photographs would be cued by the experimental protocol. In this way, participants would not be able to select activities that they felt were interesting. For convenience and simplicity, the experimental protocol was based on time; by asking people to make a record using time of day, rather than activity, it was hoped that we would be able to collect a richer, more representative set of examples of everyday activity. For example, given the instruction to record ‘everyday activity’, people may have concentrated on the more interesting aspects of their day that could have unduly skewed the recordings. Obviously, participants were granted some lee-way in what to record, e.g. they were allowed to exclude ‘personal’ activities from the study providing they recorded activity within the window of around 15 min. It is worth noting that, in this study, ‘time of day’ functioned solely as a cue to performing the act of taking a photograph (and not as an independent variable in the experiment).

2.1. Phase one: photo collection

The photo diary method involved participants taking photographs of tasks or activities that they were performing at specified times during the day.

Participants: Seven people took part in the study (3 male, 4 female: mean age 37.2 (± 19) years). The sample size was limited in order to make the subsequent analysis less onerous. The selection of participants was intended to reflect as broad a population as possible, and the occupations were, student, local government officer, teacher, social worker, gas engineer and two researchers.

Procedure: The week was split into 84, hour and a-quarter sessions (each day started at 8 am and finished at 11 pm). Twelve sessions were randomly allocated to each of the seven participants. The participants were asked to take two photographs of one activity they did during each time period. When the participants took these photographs they were asked to note the time, exposure number and a one-line

description of the activity. The objective was to get evidence of a wide range of typical everyday activities. Through this method we sought to minimize bias in the selection of activities people photographed during the time slots, and to ensure that the activities were evenly distributed over the week. In addition the participants were not told how to photograph their activity, only that the photos should be self-explanatory in describing the activity. Each of the time slots was only used once. Repetition of time slot by different photograph takers was not necessary. As mentioned above, we are not so much interested in activities at certain times, the time slots are used as a way of forcing photograph taker to get a wide range of activities and not just one which they feel are interesting.

2.2. Phase two: photo sorting

Once the activities had been recorded the sorting of the photos followed a variation on the Card Sorting Method (Maiden and Hare, 1998; Sinclair, 1995). This involved laying out all of the photographs and then searching for features that would allow photographs to be grouped. This process was conducted by a panel of judges, who performed the card sort task twice. The reason for performing the task twice was to explore variations in sorting, which would emphasize different features in the photographs.

Participants: Seven judges completed the photo-sorting task. All the participants were engineering students (4 male, 3 female, mean age 25 (± 3) years) and had not completed the photo collection task.

Procedure: The photographs were mounted on pieces of card in their pairs. Each judge was given identical verbal instructions to sort the photo cards into similar groups. No definition of 'similar groups' was given; this was to be defined by the judge. On the back of each photograph was written the time they were taken, the one line description of the activity and the user number. The judges were instructed to only look at these details if they could not ascertain the activity from the photo itself. The judges were then asked to give a heading to depict each of their sets and describe the main features that defined membership to the group. These features were called the Key Features.

One week after completing the first sorting exercise the judges were asked to resort the cards, ensuring that their classification was different to that in their first sort. This was done to increase detail and variety of the Key Features elicited by the technique. It was obvious from the first sort that judges sorted in a similar manner, with body posture and movement and location being the most popular Key Features. It is, perhaps, interesting to note how 'location' serves as a 'common-sense' descriptor of context and has been the basis of much of context-awareness research. The body posture and movement is, we felt interesting, in that the emphasis of the study has been on human performance and this implies (we feel) that 'location' was also related to human activity as opposed to being a parameter that related to a specific device that could be carried by a person. Asking more people to sort the card from the random starting position (as in sort 1) would probably have produced more of the same Key Features. By getting the same people, who had familiarity with the

photos and the features to sort again, but in a different way produced additional features. Since the desired outcome of the exercise was to produce as many and as diverse range of features as possible, the second sorting sessions went a long way to achieving this. Again the heading, features and card numbers were recorded for each group.

2.3. Phase three: context identifier sort

Each Context Identifier from the photograph sort was then written onto a post-it note. The post-it notes were then subjected to a further sort. The Post-it sort had two purposes; firstly to identify sets of similar Context Identifier identified by the judges. Secondly to give an importance weighting to each of the Context Identifiers found, allowing a designer to select the most appropriate Context Identifier for their context-aware device. The Context Identifiers do not tell a designer anything about what types of sensors to use, just the types of entities to sense.

Participants: in the final sorting phase, the authors performed the task.

Procedure: All of the elicited Context Identifiers from both photo sorts were recorded on small yellow post-it notes. These post-its were then stuck to a large piece of paper and organized into groups of connected features (i.e. ones that contained similar key words or themes). This sorting produced a concise set of theme areas or Context Identifiers, these included Location, Body Position, Objects and others. In Location you would find terms such as home, work, office, indoors, outdoors. Location was then split into subcategories such as known-locations, indoors/outdoors, and novel-locations. The final groups of Post-its can be seen below in Fig. 1. The miscellaneous category includes terms that were too general to fit into any other category; generally, they were terms that described an activity e.g. shopping, which is the activity rather than a component of the activity.

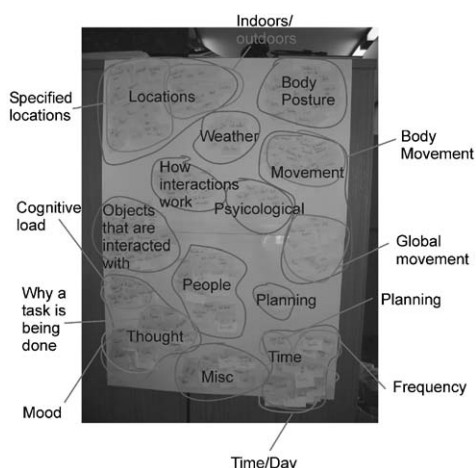


Fig. 1.

2.4. Conclusions

Study one covered a large number of everyday activities, both work and leisure orientated. By randomly allocating time slots, many mundane activities were included that might have been missed. The Context Identifiers are shown below in Table 2 in order of importance (highest first) with examples of states. The ranking was defined by the percentage of Key Features that were included in a specific Context Identifier, i.e. given that participants identified some 400 Key Features, the top Context Identifier, ‘Body Posture and Movement’, had around 96 Key Features associated with it. It is hypothesized that the highest weighted Context Identifiers are most important as they occur most often in everyday activities. Body Posture and Movement could be important because it is featured in almost all of the activities analysed (which, it must be noted, featured people performing everyday activities) whereas Weather was only a Context Identifier in a few of the activities (Fig. 2).

It is proposed that Table 2 could support a designer in developing context-aware devices. This can be achieved by using this list to determine the primary Context

Table 2
Ranked list of context identifiers

context identifier	% of total	Possible states of the context identifier
Posture and movement	24.25	Is the person moving, walking, sitting, standing?
Location	14.75	Is the location known?
Object	14.5	What object is being interacted with?
People	8.00	How many/which people are involved?
Time	7.50	What time or day is it?
Mood	6.25	Is the user relaxed/stressed?
Indoors/outdoors	5.75	Is the person indoors?
Type of interaction	5.25	How did the interaction between people of object take place?
Aim of task	4.00	Is the task being done for enjoyment/work?
Physiological indicators	3.25	Is the heart rate high?
Cognitive load	3.00	How mentally difficult is the task?
Weather	1.75	Good/bad weather?
Frequency	0.75	How frequently is the task done?
Time critical?	0.75	Is the task length (time) constrained?
Planning?	0.50	Is the task planned/spontaneous?



Fig. 2.

Identifier for an application and then work down the list adding more Context Identifiers as appropriate. The application of percentages to the Context Identifiers should not be read in any absolute term, but provide an interesting view of the relative importance of these Context Identifiers for the everyday activities captured and analysed in study one. For example, the top three Context Identifiers account for over 50% of the responses, whilst the bottom six account for only 10%. This is not to say that one should not use the Context Identifiers at the bottom of the list; one of the characteristics of ‘context’ is, of course, that the relevance of Context Identifiers will be dependent on what activities are being examined (so, for instance, a medical application or a study of mental workload might place higher ranking on physiological indicators or cognitive load than is shown in Table 2). However, if researchers wish to consider Context Identifiers in specific application domains, then the methodology presented in study one provides a relatively quick and simple means of defining this information. For the purposes of this paper, it is proposed that a wearable computer that uses context-awareness to support everyday activity ought to be designed to handle location, and posture and movement of its wearer.

3. Study two: using location to define context

3.1. *Developing a wearable context-aware system*

Having established the importance of specific Context Identifiers, it was important to determine whether context-awareness could actually be shown to be beneficial for wearable computing. After all, if it could not be demonstrated that context-awareness leads to measurable benefits to the user, why should such systems be developed? Therefore, two studies were carried out to investigate the effect of using specific Context Identifiers on user performance. In the following section, the platform, i.e. the wearable computer and the software, used in the first study will be described.

The Photo Diary study suggested that the top two Context Identifiers were body position and movement, and location. As discussed in the Introduction, Location has long been a favourite Context Identifier in the Context-awareness literature. If it could be demonstrated that Location could enhance user performance when using a context-aware device, then it is possible to propose that the other Context Identifiers might also be beneficial. Therefore, sensing Location would seem to be a sensible place to start. Thus, a system was developed that senses a user’s location and reacts to it. This system is the χ^3 .

3.1.1. *The χ^3*

The χ^3 is a wearable computer developed at the University of Birmingham.¹ The χ^3 uses a PC104 embedded PC board, and has SVGA out, two communications ports, on-board LAN and USB sound. The main unit is a 166 MHz Pentium class

¹<http://www.wear-it.net>

chipset. A MicroOptical head-mounted display is used (with its own power source and data converter), with the addition of a SVGA to NTSC converter allowing the screen to be made larger for reading text from webpages. A Garmin GPS is used for tracking the users' location. The GPS is accurate to a few metres, although can be affected by reflections from buildings.

The processor runs Windows 98. This offers the capability to run commercially available software and to share files between different computers easily. However, there is the assumption that Windows is 'power hungry'. We have found that it is possible to modify BIOS settings in such a way as to significantly reduce power requirements and to extend battery life; we typically expect some 6–8 h of battery life (and have managed to run for 10 h on full load in laboratory settings). While these times are, perhaps, too short for commercial application (one would not want to keep running out of power towards the end of the working day), they do suggest that it is possible to control power management in windows to an acceptable degree.

Fig. 3 shows the χ^3 . It is fairly small, measuring $170 \times 40 \times 100$ mm. Even with the addition of the head-mounted display and battery the system it is still comfortable, light and easy to wear (Knight et al., 2002). The entire system is housed in a shoulder strap, as shown in Fig. 4.



Fig. 3.

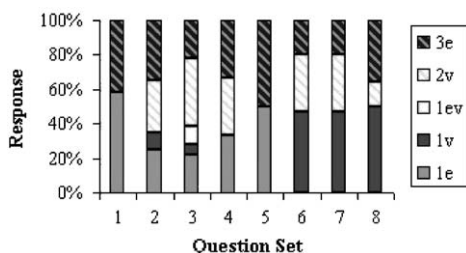


Fig. 4.

3.1.2. WECA-PC software

WECA-PC Portal is the software that analyses the data from the χ^3 Communications Port and sends a URL from an on-board database to Internet Explorer. The incoming data either arrives from a GPS or from infra-red transmitters located in buildings (although the latter option is not used in this trial). The software is written in Visual Basic. The software strips longitude and latitude coordinates from the GPS data stream. In addition to extracting coordinates, the software also performs error checking and invokes different routines depending of the absence or weakness of signals from the GPS. The coordinates are used to query a database of previously identified locations in order to call up a URL. In this application, web-pages are pre-stored on the χ^3 and we do not use the GSM link.

3.2. Objectives

In the trials presented in this paper, the aim has been to compare three conditions that reflect realistic sources of information for the task ‘find out about buildings on the University of Birmingham campus’. One source of information is the Internet; people can access web pages to find out about departments and buildings (this is Condition 2—web). A second source is the buildings themselves; people can walk around campus and look for information (this is Condition 3—world). A wearable computer ought to permit access to the Internet for people who are walking around campus and should provide the best of both worlds (this is Condition 1—wearable). Conditions 1–3 are described in more detail in Section 3.2.1.

The aim of the user trial was to determine whether a context-aware wearable computer would be useful for information retrieval tasks. It was assumed that such a device would allow users to extract information from both the World Wide Web and from the world around them. Consequently, the trial was constructed so that people would search for information in one of three conditions (web only, world only, wearable computer). In terms of Context Identifiers from study one, this study is primarily concerned with Location, i.e. through the use of GPS data.

The comparison, between world, web and wearable, was intended to test whether wearing a context-aware device led to superior performance to simply using information that was available in the world, and whether mobile access to web-pages led to superior performance to simply accessing the web on a static terminal. It was assumed that the two reference conditions, i.e. walking around and using the internet, would produce superior performance on two measures, i.e. we hypothesized that wearing the computer might impede mobility, so that the walking around condition should lead to significantly faster mean time between questions (note the mean time between questions is the average time, for the mobile conditions, walking between building and for the Internet condition the average time rested between questions); and that the internet would lead to quick information retrieval, and so should support significantly faster time to find information (particularly when one considers that the internet condition would employ the traditional mouse and keyboard and a 17" monitor familiar to all participants). There are two dependent

variables in this trial Time and Accuracy; therefore better performance would be faster and more accurate.

3.2.1. Method

Participants: 27 undergraduate students participated in the study (23 male and 4 female). Age range: 18–22. The participants were divided into three groups of 9 for this study.

Procedure: A set of questions about buildings around the university campus was devised, comprising of two questions for each building, eight questions in total (Table 3).

The participants were allocated on appearance to the base-room to one of three conditions (9 participants in each condition). The questions all had specific answers and could be answered in both conditions. The condition denoted the type of information sources available to answer all of the questions as follows:

- 1. Condition 1: Wearable Computer—the user wore the χ^3 with the WECA PC software installed and therefore had access to both environmental and virtual data. As the users approached the relevant building a single web page was displayed, in each case this page contained the answer to at least one of the questions asked about that building.
- 2. Condition 2: Internet—involved the user only having access to virtual data. The users were asked to sit at an Internet ready terminal in the School of Electronic and Electrical Engineering and only use the University of Birmingham web sites to find the answers.
- 3. Condition 3: World only—involved only environmental data. The users were asked to walk around campus with a paper map, visiting the relevant building to answer the questions.

Each group was given a brief explanation of the task and an introduction to the equipment used. Training was limited, as we wanted to see how people would cope with the novelty of the wearable computer. The questions were given to participants,

Table 3
Example questions

Location	Arrival time at building	Departure time at building	Answer	How did you get it?
<i>Library</i>				
Q1: In what year did Queen Elizabeth lay a stone in the library?			1957	Look at stone or look at picture on the web
Q2: How many shields are above the main doors?			3	http://www.is.bham.ac.uk/mainlib/about.htm or go look at the main doors

and any queries regarding task or questions addressed. Each group was then told that they had 20 min to complete the task (although they were not stopped if they went over this time).

The time taken to answer each set of questions (i.e. the time taken looking for information) and the total time completing the task were recorded. This allowed a comparison to be made of the time taken to complete each individual question, the total time taken and the time taken between questions (in Conditions 1 and 3 the time taken walking between buildings). The users answers to the questions were recorded and then marked to give them a score. Users were also asked to record whether they answered the questions making use of virtual or environmental information and any problems they encountered. At the end of the experiment, the answers were checked and the participants debriefed. A full set of results was made available to participants the week after the experiment.

3.2.2. Results

The results are divided into two parts: the first considers the effect of the different conditions on performance time, and the second considers the effect of the different conditions on participants' ability to answer the questions.

Performance time data: Table 4 shows the overall time spent answering the questions, between questions and overall task time for each condition.

A non-parametric Kruskal–Wallis test (assuming that the sample size would not produce normally distributed data) was carried out to compare the mean time spent answering questions and the results show a significant effect of condition [$\chi^2 = 11.779$; $df = 2$; $p < 0.005$]. Looking at the mean times it is apparent that the wearable computer condition (Condition 1) performed significantly faster than the other conditions.

In addition to considering how quickly participants could answer the questions, the time to complete the study and the time between attempting each question were measured. For the 'outdoor' conditions, i.e. Condition 1: Wearable and Condition 3: World, this was the time taken to complete the task at one building and walk to the next building, and Condition 2: Internet, this was the time to move to the relevant web-page. Kruskal–Wallis tests were applied to the time between questions and the total time taken to complete the task by each condition. For the 'Time between Questions' a significant difference between conditions was revealed [$\chi^2 = 6.11$; $df = 2$;

Table 4
Mean performance times for each condition

	Mean time (min)		
	To answer all questions	Between all questions	Task completion
1. Wearable	7.51	11.17	19.09
2. Internet	23.13	3.33	26.47
3. World	12.34	10.51	23.17

$p < 0.05$]. The time between questions is much greater for Condition 1: Wearable and Condition 3: World (walking around), than for the Condition 2: Internet. The difference between the total times taken to complete the test were not shown to be significant although the participants in condition 1 were seen to complete the task in the shortest amount of time.

Ability to answer questions: Table 5 compares the mean performance on question answering in terms of percentage of questions answered correctly.

A Kruskal–Wallis test indicated a significant difference in performance between conditions [$\chi^2 = 14.853$; $df = 2$; $p < 0.001$]. Inspection of Table 5 suggests that while the Condition 2: Internet and Condition 3: World performed at similar levels, the wearable condition exhibited superior performance.

Fig. 5 shows how participants in the different conditions relied on different sources of information to answer the questions. The sources of information could be the Environment (e) in which the person was walking or the Virtual display (v) of information on the World Wide Web, or could be answered using either Environment or Virtual (ve). The numbers, in Fig. 5, refer to the three experimental conditions, i.e. 1 = wearable, 2 = internet; 3 = walking around. Thus, Condition 2 only made use of virtual information (hence, 2v in Fig. 5), and Condition 3 only made use of the environment (hence, 3e in Fig. 5), but Condition 1 could conceivably use any source, i.e. 1e, 1v, 1ve. It is apparent, that questions 1 and 5 could be most easily answered using environmental information, whereas the other questions could be answered using either environmental or virtual information. What is interesting is

Table 5
Mean performance on question answering

	Condition		
	1. Wearable	2. Internet	3. World
Mean score on questions	100%	71.2%	80.5%



Fig. 5.

that the participants in Condition 1: wearable seemed to use the information that most easily provided the answer, with a possible trend to use environmental information if readily available and then resort to virtual information (unless virtual information would more easily provide the answer). For example when the participants in the Internet condition answered a question incorrectly, the participants in the Wearable condition typically elected to gain information from the environment and not the Internet. Participants in the Wearable condition successfully rejected the source that had failed to provide the answer quickly and accurately and hence maintained greater accuracy and speed of response. It can be seen that the number of users in the Wearable condition answering correctly matches the number from either the Internet or World conditions, i.e. the same information source has been used to answer the questions correctly.

3.2.3. Conclusions

The results indicate that using the wearable computer (Condition 1) not only reduced the mean question–answering time but also enabled the participants to answer all of the questions correctly. Those using this system performed more effectively and efficiently. In terms of answering questions, the mean number answered correctly was greater for the Wearable and World condition than the Internet condition (Condition 2). Participants in the World condition scored higher than expected because they made use of ‘other’ information sources in the real world, e.g. one of the questions related to degree programmes and participants in this condition all went into the relevant Undergraduate office to ask the secretary, these incidents were isolated.

It is proposed that the participants’ ability to answer each question was dependent on the source of information available to the participant. In other words there is a relationship between the source of the information (whether it is environmental or virtual) and the accuracy and speed of response for each of the specific questions. Certain questions were more easily answered based on the source of the information available. This is indicated by the results, as it is clear that the participants in the Internet and World conditions answered different questions correctly. It is further supported by the link to the choice of information source selected by the Wearable condition; the users in the Wearable condition consistently chose the most appropriate source.

Benefits of experimental design: It is important to note that if the experiment only looked at total time to complete the task, no useful results would be found. This is often the case in comparative evaluation of products and technologies; differences lie in the process rather than the outcome of task performance. The fact that times were looked at for each individual component has been useful in ascertaining the full implications of using χ^3 . The mean time taken between questions (or physically the time taken walking between buildings) is obviously greater for the wearable and walking around conditions, where walking was necessary, than for Condition 2 where the users has only to sit at a computer. What is interesting (and against our expectations) is that these times are comparable for the Wearable and World

condition. This was taken to imply that the wearable did not significantly impede performance.

Specific problems: Some users found a problem with the GPS, particularly when near a specific building. This can be accounted for by two factors. Firstly if a user is standing near the edge of the active area (the area in which WECA PC Portal will react and load the relevant web page) around each building, the GPS has a certain drift factor, this factor can then virtually move a user in and out of the active area, thus loading and unloading the web page and causing confusion. Secondly, if the area around the building has a number of high buildings, as the users walked through small gaps between the buildings the GPS may have been affected by the buildings thus confusing the system. Users also complained about the Head-Mounted Display, feeling that the resolution and size of the display make it almost impossible to read text from a page without zooming and scrolling around the screen for the relevant piece of text. However, it was not very easy to operate and many users found it frustrating. Despite the problem of reading the display, participants' performance did not appear to be unduly impaired.

4. Study three: using location and body position to define context

Study two indicated that a location-based context-aware wearable system can improve a users performance in a given information retrieval task. Study one suggested a set of Context Identifiers that could be used to find a users context. Whilst the Context Identifier of Location featured in the top three from study one, and was used in study two, posture and movement position was identified as the most important feature of context in study one. Therefore, it is suggested that this is next entity to sense. The aim of study three was to determine whether a level-two context-aware wearable computer would be useful for information retrieval tasks. It is hypothesized that adding a second level of context will perform better than an improved version of the single level used in study two.

For study three two versions of a wearable computer called the χ^{3+} were produced. System 1 took the concepts of best performing system from study two and improved it in terms of better underlying technology, for example a faster processor. The underlying concept of just sensing location as the only Context Identifier has not changed. Thus, context-awareness using Location represented the control condition of study three. System 2 was created using the same technology but with the addition of the ability to sense the users body position. These two systems can then be compared. In terms of the Context Identifiers defined in study one, study three employed Location and Body Position.

4.1. The χ^{3+}

The χ^{3+} is an upgraded version of the χ^3 used in study two. The main board is still a PC104, however it is now a 700 MHz 256 Mb RAM. In addition the USB sound card and headset that was not operational is now fully working due to a new more

Table 6
Comparison of χ^3 versions

Feature	χ^3	χ^{3+}
Main board	PC104	PC104
Processing power	166 MHz	700 MHz
RAM	32 Mb	256 Mb
Context sensors	GPS	SYSTEM ONE GPS ONLY SYSTEM TWO GPS & Accelerometer
Carrying system	Across shoulder system	'Camel Back' system
Head mounted display	Micro-Optical, single eye, 200×300 resolution, translucent	Cy-Visor, dual eye, 800×600 resolution, opaque
Sound		USB headphones and microphone
Power supply	5A	8A
Operating system	Windows [®] 98	Windows [®] 2000
Control device	Thumb mouse	Thumb mouse
Internet connectivity		GSM Nokia 6210 Mobile phone, 9.2 kbps data rate

powerful power supply and different chip set, and the GSM Internet connection is working. The Micro-Optical head-mounted display has been replaced by a Personal Display Systems Cy-Visor, which has display giving a true 600×800 resolution making text clearer and more readable. However, this display was not translucent and therefore did make walking more difficult. As with study one a Garmin GPS was used, for the second level of context a two-axis leg accelerometer was added. Fig. 5 shows participants wearing this kit. Table 6 shows a comparison of the χ^3 and the χ^{3+} .

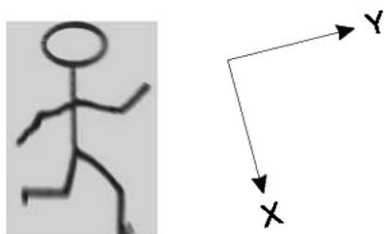
4.2. SSW—stand, sit, walk software

The WECA PC Software used in study one was replaced with SSW Software. SSW captures and analyses data from the GPS and the accelerometer. Data from the accelerometer is used to tell whether the user is sitting, standing or walking. The accelerometer has an *X*- and *Y*-axis. When the accelerometer is attached to the leg the *X*-axis runs parallel to the leg and the *Y*-axis at right angles to it. SSW takes an average reading over 2 s (approximately 300 values) to get a value of the acceleration that checks whether the user is moving or not. Once SSW has decided the users Body Position it queries a database with the GPS coordinates and Body Position and retrieves a URL associated with the context. In study two the URL was passed to Internet Explorer however SSW has a custom Internet browser built in so the URL is sent to this. The browser can be seen in Fig. 6. The user is told the Location and Body Position via the headset; they are also shown at the top of the browser.

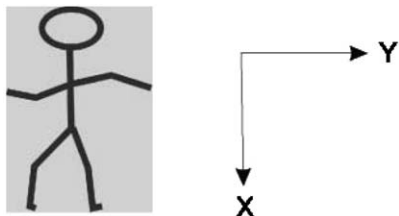
When the user is walking there is a much higher acceleration in both directions, particularly the *X*-axis. Thus SSW tests when the average *X* value is greater than 1.6g and declares that the user must be walking



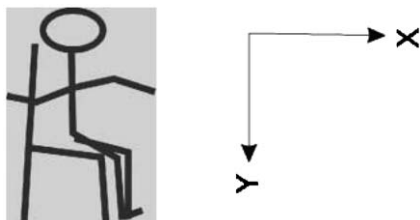
Fig. 6.



When the user is in the standing position, gravity acts on the X -axis, making the acceleration $1g$ or greater. In addition since the user is stationary the Y acceleration is small. Therefore, SSW tests when the average X near zero is greater than or equal to 1 and tests when the average Y value is small (less than $0.5g$) and declares this the standing position



When the user is in the sitting position, gravity acts on the Y -axis, making the acceleration $1g$ or greater. In addition since the user is stationary the X acceleration is near zero. Therefore, SSW tests when the average Y value is greater than or equal to 1 and tests when the average X value is small (less than $0.5g$) and declares this the sitting position



4.3. Objectives

In terms of the experiment (method and conditions explained in the next section), in System 1 used in Condition 1 (Location Only) the accelerometer is switched off and the software behaves similarly to wearable in study one, a web page associated with the current location is shown. In System 2 used in Condition 2 (Location and Body Position) however the software behaves differently. If the user is walking a blank screen is shown (the user does not therefore need to look at the screen, which should make it easier to walk), when the user approaches a building of interest SSW tells the user they are walking near the building via audio. They are also told if they want more detailed information to stand still or sit down. If the user sits down the web page is shown in full (the same web page as when the user approaches the building in Condition 1). However, if the user stands still they are shown a cut down version of the web page. At present the editing of the web page is pre-done manually, but it is proposed that this could be automated. Thus, the information displayed to the user is adapted for their current context. For example, when the user is walking, the information presented needs to be simple and non-distracting. When a user stands still they can cope with more detailed information; they can devote more attention to the task rather than looking where they are going. Again if the user sits down they can devote still further attention to the information being presented to them and cope with a higher level of information. In terms of study three, it is hypothesized that Condition 2 (Location and Body Position) will perform better since less searching is needed since the information is automatically searched/edited by the system and presented in a way that is more helpful to the user in their given context.

4.4. Method

Participants: 10 undergraduate students participated in the study (9 male and 1 female). Age range: 18–22. Participants were divided into two groups of 5 for this study.

Procedure: A set of questions about buildings around the university campus was devised, comprising of two questions for each building, six questions in total. The participants were allocated on presentation to the base-room to one of two conditions (five participants in each condition). The condition denoted the type of information source to be used to answer all of the questions as follows:

1. Condition 1: (Location Only) Wearable Computer with level-one context—The user wore the χ^3 with SSW Software (see below) installed, with the GPS attached but without the accelerometer attached. Therefore Location only was sensed.
2. Condition 2: (Location and Body Position) Wearable Computer with level-two context—The user wore the χ^3 with the SSW Software installed with the accelerometer and GPS attached. Therefore, Location and Body Position were sensed.

Each group was given a brief explanation of the task and an introduction to the equipment used. Training was limited, as we wanted to see how people would cope with the novelty of the wearable computer. The questions were given to participants, and any queries regarding task or questions addressed. Each group was then told that they had 30 min to complete the task (although they were not stopped if they went over this time).

The time taken to answer each set of questions was recorded (i.e. the time taken looking for information). This allowed comparison to be made of the time taken to complete each individual question. Users were asked to record whether they answered the questions making use of virtual or environmental information and any problems they encountered. At the end of the experiment, the answers were checked and the participants debriefed. A full set of results was made available to participants the week after the experiment.

4.5. Results

The results are divided into two parts: the first considers the effect of the different conditions on performance time, and the second considers the effect of the different conditions on participants' ability to answer the questions.

Performance time data: The mean time taken to answer the questions has greatly decreased from 20:12 in Condition 1 (Location Only) to 8:00 in Condition 2 (Location and Body Position).

If the mean times to answer the questions in condition one and condition two are compared, a difference of 12 min can be seen. It was felt that most of this time difference was due to the amount of information being downloaded and not due to the performance of the user. Therefore, an additional five users were asked to complete an altered Condition 1 on a computer in the laboratory using a high-speed line thus reducing the download time. Participants were made to use the same thumb mouse and headset to simulate Condition 1. The mean time to answer the questions taken by the altered Condition 1 was then subtracted from the mean time taken to answer the questions in Condition 1 giving an estimate of the mean time spent downloading information (12 min, 13 s). The mean download time was then divided by the number of questions giving a mean download time per question of 2 min, 2 s. The mean time taken to answer the questions and the mean total time taken to complete the task were then adjusted to take account of the download time. For both Conditions 1 and 2 the mean download time per question was multiplied by the number of questions that required information to be downloaded (In Condition 1 all questions required downloading of information, but in Condition 2 fewer questions required downloading since the some of the information was stored on the hard drive, in most cases only 2 or 3 questions required downloading, depending on the way the users answered the questions.) was subtracted from each users time taken to answer the questions and total time taken to complete the task, then means were then recalculated and are shown in [Table 7](#). Thus, a fairer result is shown.

After adjusting for the downloading time a reduction in the mean time taken to answer the questions and the total time to complete the task can be seen in [Table 7](#).

Table 7

Mean time taken to answer questions and to complete the task adjusted for the time taken to down load pages from the Internet

	Download adjusted	
	Condition 1: location only DA	Condition 2: location and body DA
Mean time answering questions	07:59	03:56
Total time	15:47	10:20
Standard deviation time answering questions	01:47	02:33
Standard deviation total time	02:21	01:31

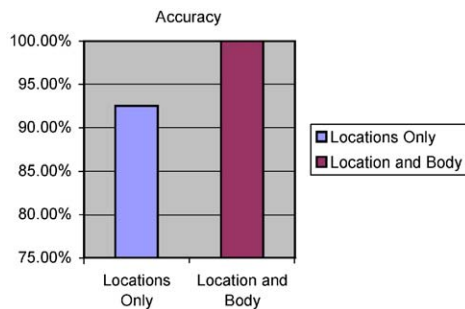


Fig. 7.

Kruskal–Wallis tests were carried out on the download-adjusted times, these show that both times give significant results, mean time answering questions [$\chi^2 = 3.962$; $df=1$; $p<0.05$] and Total time to complete task [$\chi^2 = 6.902$; $df=1$; $p<0.05$]. Thus, condition two performed better than condition one.

Ability to answer questions: From Fig. 7, it can be seen that Condition 2 performed 7.5% more accurately than Condition 1. This result is not statistically significant.

4.6. Conclusions

From the results of Study 2, it can be concluded that adding a second Context Identifier increased the performance of a user completing a given task, particularly in terms of time (although the accuracy data have not seen much improvement between conditions). The mean time taken to answer the questions and the mean total time to complete the task has been significantly reduced by the addition of a second context level. The result should not be taken simply to indicate that more Context Identifiers lead to better performance. Rather, each Context Identifier has a positive effect on a specific aspect of performance, i.e. Location provides a direct route to information that is relevant to a particular place, and Body Position provides a means of managing the presentation of the information, depending on what the user is doing.

Specific problems: Since the HMD was not translucent, and even though it could be pushed up and down, some people found it difficult to walk at a normal pace. A

second problem was the slow download speed of the mobile phone; this has been accounted for by the download adjustment and therefore should not influence the results. In the future the campus may be covered by high-speed wireless local area network (IEEE 802.11b) coverage, which would remove this problem.

5. Discussion

In conclusion, the Photo Diary Method has allowed us to define that Context Identifiers that are important in defining a users context. Designers could use the list of Context Identifiers, presented in Table 2, and the links between them to select what to sense in a given situation. Work must now be carried out to find technological means of sensing each of the Identifiers.

The user trials involving χ^3 and χ^{3+} have answered the question: ‘does context-awareness support user performance?’ The results indicate that context-awareness can improve user performance on information retrieval tasks. In fact, a considerable improvement in user task proficiency can be seen when using the wearable system.

It is proposed that there are two main conclusions that can be drawn from the user trial results. The first is that providing people with appropriate information ‘just-in-time’ can be beneficial to the performance of specific tasks. This is not too surprising and basically offers empirical support to some of the underlying assumptions that the ‘context-aware’ world has been using for some time. The second is that, given the option to use more than one information source, people are good at selecting the most appropriate source. In other words, the results indicate that context-aware technology is not simply about providing the information, but about providing a source of information to which users can flexibly respond.

Acknowledgements

This work was partly supported by European Union grant IST-2000-25076 ‘Lab of Tomorrow’ and EPSRC Grant GR/R33830.

References

- Abowd, G.D., Mynatt, E.D., 2002. Charting past, present, and future research in ubiquitous computing. In: Carroll, J.M. (Ed.), *Human–Computer Interaction in the New Millennium*. Addison-Wesley, Boston, MA, pp. 513–536.
- Ashbrook, D., Starner, T., 2002. Learning significant locations and predicting user movement with GPS. Digest of Papers of the Sixth International Symposium on Wearable Computers. IEEE Computer Society, Los Alamitos, CA, pp. 101–108.
- Asthana, A., Cravatts, C., Krzyzanowski, P., 1994. An indoor wireless system for personalized shopping assistance. Proceedings of the IEEE Workshop on Mobile Computing Systems and Applications, Santa Cruz, CA, December 8–9.
- Baber, C., 2001. Wearable computers: a human factors review. *International Journal of Human–Computer Interaction* 13 (2), 123–145.

- Baber, C., Haniff, D.J., Woolley, S.I., 1999. Contrasting paradigms for the development of wearable computers. *IBM Systems Journal* 38 (4), 551–565.
- Brown, B.A.T., Sellen, A.J., O'hara, K.P., 2000. A diary study of information capture in working life. In: Turner, T., Szwillus, G., Czerwinski, M., Paterno, F. (Eds.), *CHI 2000*. ACM Press, New York, pp. 438–445.
- Eldridge, M., Lamming, M., Flynn, M., 1992. Does a video library help recall? In: Monk, A., Diaper, D., Harrison, M.D. (Eds.), *People and Computers VII*. Cambridge University Press, Cambridge, pp. 257–269.
- Farrington, J., Moore, A.J., Tilbury, N., Church, J., Biemond, P.D., 1999. Wearable sensor badge and sensor jacket for contextual awareness. *Digest of Papers of the Third International Symposium on Wearable Computers*. IEEE Computer Society, Los Alamitos, CA, pp. 107–113.
- Feiner, S., MacIntyre, B., Hollerer, T., Webster, A., 1997. A touring machine: prototyping 3D mobile augmented reality systems for exploring the urban environment. *Digest of Papers of the First International Symposium on Wearable Computers*. IEEE Computer Society, Los Alamitos, CA, pp. 176–177.
- Knight, J.F., Baber, C., Schwirtz, A., Bristow, H., 2002. The comfort assessment of wearable computers. *Digest of Papers of the 6th International Symposium on Wearable Computing*. IEEE Computer Society, Los Alamitos, CA, pp. 65–74.
- Lamming, M., Flynn, M., 1994. Forget-me-not-intimate computing in support of human memory. *Proceedings of FRIEND21: International Symposium on Next Generation Human Interface*, Meguro Gajoen, Japan.
- Lind, E.J., Jayaraman, S., Park, S., Rajamanickam, R., Eisler, R., Burghart, G., McKee, T., 1997. A sensate liner for personal monitoring applications. *Digest of Papers of the First International Symposium on Wearable Computers*. IEEE Computer Society, Los Alamitos, CA, pp. 98–107.
- Maiden, N.A.M., Hare, M., 1998. Problem domain categories in requirements engineering. *International Journal of Human-Computer Studies* 49, 218–304.
- Marmasse, N., Schmandt, C., 2000. Location-aware information delivery with commotion. *HUC 2000*, pp. 157–171.
- Randell, C., Muller, H., 2000. The shopping jacket: wearable computing for the consumer. *Personal Technologies* 4, 241.
- Rhodes, B., 1997. The wearable remembrance agent: a system for augmented memory. *The First International Symposium on Wearable Computers*. IEEE Computer Society, Los Alamitos, CA, pp. 123–128.
- Rieman, J., 1993. The diary study: a workplace-oriented research tool to guide laboratory efforts. In: Ashlund, S., Mullet, K., Henderson, A., Hollnagel, E., White, T. (Eds.), *INTERCHI*. Addison-Wesley, Reading, MA, pp. 321–326.
- Schmidt, A., Gellerson, H.-W., Beigl, M., 1999. A wearable context-awareness component: finally a good reason to wear a tie. *Third International Symposium on Wearable Computers*. IEEE Computer Society, Los Alamitos, CA, pp. 176–177.
- Sinclair, M., 1995. Subjective assessment. In: Wilson, J.R., Corlett, E.N. (Eds.), *Evaluation of Human Work*. Taylor & Francis, London, pp. 69–100.
- Tolmie, P., Pycock, J., Diggins, T., Maclean, A., Karsenty, A., 2002. “Unremarkable Computing”, *Chi '02*, Vol. 4(1), Minneapolis, MN, USA, 20–25 April 2002, pp. 399–406.
- Want, R., Hopper, A., Falcao, V., Gibbons, J., 1992. The active badge location system. *ACM Transactions on Information Systems* 10 (1), 91–102.